PATENT

#12
3/12/04
A.W

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appl. No.	:	09/651,073	Confirmation No. 2123
Applicant	:	Reiner KRAFT et al.	
Filed	:	August 30, 2000	
TC/A.U.	:	2175	
Examiner	:	Charles RONES	
Docket No.	:	ARC9-2000-0111-US1	
Customer No.	:	23334	

APPELLANT BRIEF

Commissioner of Patents and Trademarks
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This Appellant Brief is filed in response to a Final Office Action dated September 30, 2003 and a Notice of Appeal filed December 30, 2003, the due date for response being March 1, 2004, the next business day after February 29, 2004. Reconsideration of the Application, withdrawal of the rejections and allowance of the claims is respectfully requested.

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 or facsimile transmitted to the U.S. Patent and Trademark Office on the date March 1, 2004 By: Karen Taragowski

Signature: Karen Taragowski
Applicant, Assignee, or Representative

03/09/2004 AWONDAF1 00000118 090441 09651073

01 FC:1402 330.00 DA

I. REAL PARTY IN INTEREST

The real party in interest is International Business Machines (IBM) of Armonk, NY.

II. RELATED APPEALS AND INTERFERENCES

There are no related appeals or interferences.

III. STATUS OF CLAIMS

Claims 1-26 are pending. Claims 1-26 were finally rejected in the Office Action dated September 30, 2003. Claims 1-26 are on appeal.

Attached hereto is an Appendix containing a copy of claims 1-26 (in their current form), which are the claims involved in this appeal.

IV. STATUS OF AMENDMENTS

The Examiner issued a rejection of claims 1-26 in the Office Action of March 26, 2003. An Amendment was filed on July 28, 2003 amending the claims. The Amendment was entered by the Examiner. In a Final Office Action dated September 30, 2003, the Examiner stated that the remarks and claim amendments contained in the Amendment of July 28, 2003 were not persuasive and issued a final rejection of claims 1-26.

V. SUMMARY OF THE INVENTION

The present invention provides the functionality of performing a convenient and facile comparison of similarity between documents based on a dynamic set of document identifiers. The present invention relates to the field of data processing, and particularly to a software system and associated method for

use with a search engine. The engine searches data maintained in systems that are linked together over an associated network such as the Internet. More specifically, this invention pertains to a computer software product for determining, comparing, and representing the similarity between documents using a drag and drop Graphical User Interface (GUI) within a dynamically generated list of document identifiers. The invention uses this drag and drop GUI interface for convenient selection of document identifiers for further comparison. Then processing of a similarity analysis request using a configurable similarity algorithm is executed; this processing can be done on the client, proxy or server side. When the comparison process is completed, the GUI presents the similarity result of the comparison process as a Venn Diagram to show the level of similarity between the selected documents. See Abstract of the specification of Appellant's invention.

The Appellant's invention is further directed towards a method for measuring the similarity between documents, especially documents which are returned as part of search results in a search engine. The measurement of similarity is provided as a numeric value as an indicator at to how much the content and/or the structure of two or more documents returned by the search query are similar. The present invention extracts from a search engine each of the search results using a DTD (data type descriptor) scheme. This is necessary because each search engine uses a different format for search result presentation. The present invention solves the time-consuming problem of whether a given document returned in a search result, e.g. document A, has similar content and/or structure of a second document, e.g. document B, returned as part of a search result without the need for the user to open document B. The present invention provides a numeric value as a measure of similarity between documents. Using this numeric value as a indicator, a user can quickly determine how closely related a given document A is to second document B and/or other documents. See specification of Appellant's invention at page 10, line 25-28, and page 13, lines 12-16 and lines 26-28.

Take for example, the search query of the word "patent." The word patent appears many times in the MPEP, but the search query does not indicate how similar the contents of the search result are either in content or structure. Continuing with this example, suppose a user was searching the USPTO website with the search query "35 U.S.C. § 103(a)". Many references would be found including both the MPEP and 35 U.S.C. § 103(a). Using the present invention, the user can compare the similarity of the search results. In this case the document entitled MPEP and the document 35 U.S.C. § 103(a). Continuing further, the MPEP includes the 35 U.S.C. § 103(a) so there would be a high similarity when the user compares the contents and/or structure of 35 U.S.C. § 103(a) with the MPEP. Whereas the search term "35 U.S.C. § 103(a)" by itself returns only all the documents containing this term and perhaps the number of occurrences of this term in the documents returned as part of the search result. However, the number of times a search term appears in a document is not the same as how similar two or more documents are in content and/or structure.

VI. SUMMARY OF THE CITED REFERENCES

A. Horowitz (U.S. 6,236,987)

Horowitz is directed towards an information system and method that provides organizational and navigational aids to a user to facilitate exploration and analysis of a document collection. The system includes a document collection containing a plurality of documents and a knowledge base containing a plurality of topics. Each topic expresses an idea or concept, and is associated with a set of terms which describe the topic and a set of documents in the document collection which are about the topic. Each topic also has topic-subtopic relationships with selected other topics, forming local topic hierarchies.

A query analysis module receives a current query and processes the query against the document collection to select a set of documents that satisfy the query. A dynamic content organization module processes the document set

according to defined parameters and a user selection or automatic selection of a desired topic arrangement to create various types of topic arrangements. These topic arrangements include supertopics, subtopics, perspective topic, and theme topic arrangements.

A supertopic arrangement is a set of parent topics of a topic derived from the query, which parent topics best generalize the document set. A subtopic arrangement is a set of subtopics of a topic derived from the query which best cover and partition the document set. A perspective topic arrangement has perspective topics, each of is a parent topic of a set of subtopics that cover and partition the document set. A theme topic arrangement has theme topics, each of which expresses a major subject or concept that describes the document set and distinguishes it from the rest of the document collection. See Abstract of U.S. Pat. No. 6,236,987.

B. Singhal (U.S. 6,163,782)

The Singhal reference is directed towards a method for enabling effective and efficient storage, indexing, searching and retrieval of data information from a large data text corpora. Singhal allows for data queries to be searched and the appropriate data information to be retrieved. Large global collections of data are broken down into smaller sub-collections. The sub-collections are stored independently one from the other, as in separate physical locations or simply in separate data tables within the same physical location, and can be connected one to the other through a network. As data are added to the large global collection overall, it can be sent and added to individual sub-collections and/or can be formed into a further sub-collection.

Once the individual sub-collections have been identified, each performs its own indexing function. In carrying out the indexing function, each sub-collection creates its own sub-collection view consisting of statistical information generated from what is commonly referred to as an inverted index. An inverted index is an index by individual words listing documents which contain each individual word.

The indexing function itself can be carried out in any method. Each sub-collection possesses an independent sub-collection view based upon that sub-collection's inverted index. When data information is added to the sub-collection, the indexing function is carried out again and the sub-collection's view can be re-compiled from a new inverted index.

Upon completion of each sub-collection view, the sub-collection view is sent to and/or gathered by a global collection custodian. The global collection custodian may either request from each sub-collection that it send its sub-collection view, and/or each of the sub-collections may spontaneously send the sub-collection view to the global collection custodian upon completion. Upon collection at the global collection custodian of all of the sub-collection's views, the global collection custodian builds a "global view" on the basis of the sub-collection views. Necessarily, the global view is likely to be different from each of the individual sub-collection views. Once the global view has been compiled, it is sent back to each of the sub-collections.

In this manner then, a distributed data retrieval system is built and is ready for search and retrieval operations. To search for a particular piece of data information, a system user simply enters a search query. The search query is passed to each individual sub-collection and used by each individual sub-collection to perform a search function. In performing the search function, each sub-collection uses the global view to determine search results. In this manner then, search results across each of the sub-collections will be based upon the same search criteria (i.e., the global view). The results of the search function are passed by each individual sub-collection to the global collection custodian, or the computer which initiated the search, and merged into a final global search result. The final global search result can then be presented to the system user as a complete search of all data information references. See Summary of the Invention of U.S. Pat. No. 6,163,782.

C. Goiffon (U.S. 6,453,312)

The Goiffon reference is directed towards a computer-implemented system and method for allowing users to interactively develop search queries. The system performs query development utilizing a hierarchical concept tree stored in memory, wherein the nodes of the concept tree are concepts that describe various search topics. Parent/child relationships are created between the concepts, with children concepts describing sub-categories of a parent concept, and so on. Any concept at any level in the tree structure may be related to one or more character strings descriptive of the related concept.

Query development is performed by traversing the various relationships in the hierarchical tree structure to selectively add related character strings to a potential query. The user is allowed to control the manner and extent of the traversal, and is further allowed to de-select any located concepts and character strings for further use in query development. Traversal of the tree is completed according to user specification, with only limited user intervention required to select or de-select the various located character strings for inclusion in a query string. All query string development is fully automated based on parameters specifying the extent of the concept tree traversal. After traversal of the tree structure is completed, the located character strings that remain selected based on user specifications are added to a query string.

The query string may further be modified to include logical operations. The string may be used to manually or programmatically invoke a wide variety of tools for use in performing searches of text documents, file systems, or a collection of web pages. The system allows the contents and structure of the hierarchical concept tree to be modified by the user. Both the concepts and character strings included in the hierarchical tree may be edited by the user to include search terms tailored to specific search requirements. Additionally, relationships existing between two concepts, or between a concept and a character string, may be added, deleted, or modified to thereby allow the user to control the manner in which query development proceeds. See Abstract of U.S. Pat. No. 6,453,312.

D. Baisley (U.S. 6,502,112)

The Baisley reference is directed towards a method on a computer system for processing XMI-based XML documents. The method compares two such XMI-based XML documents for identical content. The method begins with the step of parsing each of the documents to create for each a semantic graph of the document's objects. Next, a list of names of properties for each of the objects having significant order is read. For each of the objects, and then for each object's property not listed as having significant order, values of said properties are sorted. Finally, the objects of the semantic graphs are compared. See Abstract of U.S. Pat. No. 6,502,112.

E. Chu (U.S. 6,427,146)

The Chu reference is directed towards a database event detection and notification system including a rule definer for defining a high level rule which can include conceptual terms (e.g. bad, heavy) and as cooperative operators (e.g. approximate, similar-to, near-to). A rule converter converts the high level rule into a low level rule in which the conceptual terms and cooperative operators are quantified. An event manager detects and evaluates an event generated by the database or a Local Event Detector (LED).

A rule manager applies the low level rule to the event detected by the event manager, and an action manager performs an action in accordance with the application of the rule by the rule manager. The action can include notifying a specified person or program that the event has occurred. The rule converter and the action manager utilize a Type Abstraction Hierarchy (TAH) for converting the high level rule into the low level rule and performing the action respectively. The rule comprises an attribute. The TAH comprises leaves corresponding to instances of the attribute, and a hierarchical arrangement of nodes which specify ranges that include at least one of the instances respectively. The rule converter and the action manager are configured to convert the high level rule into the low

level rule and perform the action by relaxing a value of the attribute in accordance with the TAH. An existing rule can be modified or a new rule inserted into the system without shutting down the system and recompiling all the rules. See Abstract of U.S. Pat. No. 6,427,146.

VII. ISSUES

Whether claims 1, 19, 20, 22 and 23-26 are unpatentable under 35 U.S.C. §103(a) over Horowitz in view of Singhal.

Whether claim 2 is unpatentable under 35 U.S.C. §103(a) over Horowitz in view of Singhal and further in view of Goiffon.

Whether claims 3-5 are unpatentable under 35 U.S.C. §103(a) over Horowitz in view of Singhal, Goiffon and Baisley.

Whether claims 6-18 are unpatentable under 35 U.S.C. §103(a) over Horowitz in view of Singhal, Goiffon, Baisley and Chu.

Whether claim 21 is unpatentable under 35 U.S.C. §103(a) over Horowitz in view of Singhal and Chu.

VIII. GROUPING OF CLAIMS

Group I: Claims 1-18 stand or fall together.

Group II: Claims 19, 20-22, and 23-25 stand or fall together.

IX. ARGUMENT

A. WHETHER CLAIMS 1, 19, 20, 22 and 23-26 ARE UNPATENTABLE OVER HOROWITZ IN VIEW OF SINGHAL

In the Examiner's Final Office Action of September 30, 2003, the Examiner rejected claims 1, 19, 20, 22 and 23-26 under 35 U.S.C. § 103(a) as being unpatentable over Horowitz in view of Singhal. Appellant respectfully submits that claims 1, 19, 20, 22 and 23-26 under 35 U.S.C. § 103(a) are patentable over Horowitz and/or in view of Singhal because neither Horowitz, nor Singhal nor any combination of the two reference teaches the claimed limitations of

- Independent claim 1

  determining if a search engine is coupled to a comparison system for comparing content of at least two documents identified in the search results, wherein the comparison system returns a numeric similarity value which represents the similarity of the documents.

- Independent claim 19, 20, and 23

  a comparison module for comparing contents of at least two documents for similarity, wherein the comparison module returns a numeric similarity value which represents the similarity of the documents

As the Examiner states in the Final Office Action on page 3, *"Horowitz et al. discloses the claimed invention except for comparing content of at least two documents identified in the search results*" and goes on to combine Singhal.[1] To begin Horowitz as applied to Singhal is silent on "determining if a search engine is coupled to a comparison system for comparing content of at least two documents identified in the search results, wherein the comparison system returns a numeric similarity value which represents the similarity of the

[1] Appellant makes no statement whether such combination is even proper.

documents." The Examiner cites on page 3 of the Final Office Action Singhal at col. 6, lines 48-67 and col. 7, lines 1-11. The cited section of Singhal is reproduced for convenience below.

> *Having indexed the search query, a simple formula is used to assign a numeric score to every document retrieved in response to the search query. This simple formula, referred to as a "vector inner-product similarity" formula can be as follows:*
>
> *Document score = summation from I to n of (weight.sub.i,Query x weight.sub.i,Doc)**2*
>
> *where weight.sub.i,Query is the weight of word.sub.i in the search query and weight.sub.i,Doc is the weight of wordi in the document being scored. Each document is then sent to the central computer 310, via communication paths 4.1, from the local computer nodes 320, 330 and 340.*
>
> *In step 500 of FIG. 1, once all search results have been returned to the central computer via communication paths 4.1, the central computer 310 merges the variously retrieved documents into a list by comparing the numeric scores for each of the documents. The scores can simply be compared one against the other and merged into a single list of retrieved documents because each of the local computers 320, 330 and 340 used the same Global View 510 for their search process. Upon completion of the merging of the documents, a complete list is presented to the system user. How many of the documents are returned to the user can, of course, be pre-set according to user or system criteria. In this manner then,*

[2] The equation is abbreviated in this section because the mathematical symbol is readily copyable from the USPTO patent database.

only the documents most likely to be useful, determined as a result of the system user's search query entered, are presented to the system user.

Singhal is clearly teaching how a numeric score is assigned to each document returned in response to a search query. The Appellant respectfully submits that how often the words of a "search query" appears in a document as part of a search result, is not the same as measuring the similarity in content and/or structure of two or more documents which are returned from a search query. In the first instance, the occurrence the words of a search query in a document is returned. In the second instance an indication of how similar the contents and/or structure of two or more documents is returned. The contents of a document includes more than just the "search query" and the similarity of two documents therefore is more than a measure of the occurrence of the search query alone. Singhal is silent on comparing the contents of the documents themselves and instead teaches returning a document score for each of the documents retrieved as compared to the initial search query. Accordingly, the independent claims 1, 19, 20, and 23 distinguish over Horowitz taken alone and/or in view of Singhal for at least this reason.

Moreover, the Federal Circuit has consistently held that when a §103 rejection is based upon a modification of a reference that destroys the intent, purpose or function of the invention disclosed in the reference, such a proposed modification is not proper and the *prima facie* case of obviousness cannot be properly made. See In re Gordon, 733 F.2d 900, 221 USPQ 1125 (Fed. Cir. 1984). Here the intent, purpose and function of Singhal is to generate "document score" based on the response to a "search query" using a "vector inner product similarity." In contrast, the intent and purpose of the present invention is to generate "a numeric similarity value which represents the similarity of the documents" themselves with respect to each other, where the documents are returned from a search result. The present invention performs a similarity comparison between two or more documents after the search results are

returned. In contrast, Singhal is teaching measuring similarity based only on the search query string itself. Not only is Singhal teaching similarity of the search string to each document separately but this is not performed after the search results are returned but rather during the search process itself. The Appellant respectfully submits that the measurement of similarity based on "search query" is very different than the "similarity of documents" themselves, which are returned as a result of a search. The Appellant respectfully submits that modifying Singhal to provide a "numeric similarity" between two or more documents in the search results destroys Singhal's intent on providing similarity based on the "search query" as opposed to the similarity of the contents and/or structure of the documents themselves. Accordingly, the present invention is distinguishable over Horowitz taken alone and/or in view of Singhal for this reason as well.

Independent claims 1, 19, 20, and 23 distinguish over Horowitz taken alone and/or in view of Singhal. All the remaining claims in this section, i.e. dependent claims 22 and 24-26, depend from independent claims 20, and 23, respectively. Since dependent claims contain all the limitations of the independent claims 22 and 24-26 distinguish over Horowitz taken alone and/or in view of Singhal, as well.

B. WHETHER CLAIM 2 IS UNPATENTABLE OVER HOROWITZ IN VIEW OF SINGHAL AND GOIFFON

As discussed above, independent claim 1 distinguishes over the Horowitz reference, the Singhal reference, and the combination of the two references. The Horowitz reference does not by itself, or in combination with the Singhal reference, teach, anticipate, or suggest all of the recited elements of independent claim 1. Thus, all dependent claims that depend from independent claim 1, respectively, also distinguish over the Horowitz reference, the Singhal reference, and the combination of the two references. Therefore, dependent claim 2, which depends from independent claim 1, also distinguishes over the Horowitz reference and the Singhal reference. Moreover, since the present invention has

been distinguished from the Horowitz reference and the Singhal reference, the present invention has further been distinguished from any further combinations with these two references, including a combination including the Gioffon reference.

C. WHETHER CLAIMS 3-5 ARE UNPATENTABLE OVER HOROWITZ IN VIEW OF SINGHAL, GIOFFON, AND BAISLEY

As discussed above, independent claim 1 distinguishes over the Horowitz reference, the Singhal reference, and the combination of the two references. The Horowitz reference does not by itself, or in combination with the Singhal reference, teach, anticipate, or suggest all of the recited elements of independent claim 1. Thus, all dependent claims that depend from independent claim 1, respectively, also distinguish over the Horowitz reference, the Singhal reference, and the combination of the two references. Therefore, dependent claims 3-5, which depend from independent claim 1, also distinguish over the Horowitz reference and the Singhal reference. Moreover, since the present invention has been distinguished from the Horowitz reference and the Singhal reference, the present invention has further been distinguished from any further combinations with these two references, including a combination including the Gioffon reference and the Baisley reference.

D. WHETHER CLAIMS 6-18 ARE UNPATENTABLE OVER HOROWITZ IN VIEW OF SINGHAL, GIOFFON, BAISLEY AND CHU

As discussed above, independent claim 1 distinguishes over the Horowitz reference, the Singhal reference, and the combination of the two references. The Horowitz reference does not by itself, or in combination with the Singhal reference, teach, anticipate, or suggest all of the recited elements of independent claim 1. Thus, all dependent claims that depend from independent claim 1, respectively, also distinguish over the Horowitz reference, the Singhal reference,

and the combination of the two references. Therefore, dependent claims 6-18, which depend from independent claim 1, also distinguish over the Horowitz reference and the Singhal reference. Moreover, since the present invention has been distinguished from the Horowitz reference and the Singhal reference, the present invention has further been distinguished from any further combinations with these two references, including a combination including the Gioffon reference, the Baisley reference and the Chu reference.

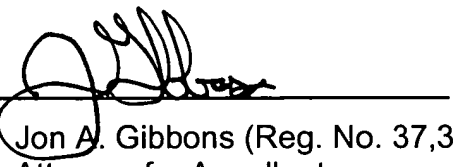
E. WHETHER CLAIM 21 IS UNPATENTABLE OVER HOROWITZ IN VIEW OF SINGHAL AND CHU

As discussed above, independent claim 20 distinguishes over the Horowitz reference, the Singhal reference, and the combination of the two references. The Horowitz reference does not by itself, or in combination with the Singhal reference, teach, anticipate, or suggest all of the recited elements of independent claim 20. Thus, all dependent claims that depend from independent claim 20, respectively, also distinguish over the Horowitz reference, the Singhal reference, and the combination of the two references. Therefore, dependent claim 21, which depends from independent claim 20, also distinguishes over the Horowitz reference and the Singhal reference. Moreover, since the present invention has been distinguished from the Horowitz reference and the Singhal reference, the present invention has further been distinguished from any further combinations with these two references, including a combination including the Chu reference.

X. CONCLUSION

For the reasons stated above, Appellant respectfully contends that each claim is patentable. Therefore, reversal of all rejections is courteously solicited.

Respectfully submitted,

Dated: March 1, 2004 By: 

Jon A. Gibbons (Reg. No. 37,333)
Attorney for Appellant
Fleit, Kain, Gibbons, Gutman & Bongini
One Boca Commerce Center, Suite 111
551 N.W. 77th Street
Boca Raton, FL 33487
Tel. (561) 989-9811
Fax (561) 989-9812

PLEASE Direct All Correspondence to Customer Number **23334**

APPENDIX

1. A method for comparison of documents found on a network interconnected with a plurality of information processing units and hub processing units, the method on an information processing unit comprising the steps of:

receiving a user search request on a concept of interest to a user;

returning search result items based upon the user search request; and

determining if a search engine is coupled to a comparison system for comparing content of at least two documents identified in the search results, wherein the comparison system returns a numeric similarity value which represents the similarity of the documents.

2. The method as defined in claim 1, wherein if the determining step determines that the search engine is supported by the comparison system, then:

parsing the search result items by a result set manager; and

identifying any document identifiers in the search result items and marking them by the result set manager.

3. The method as defined in claim 2, further comprising the step of:

retrieving knowledge of a structure and content of the search result items by the result set manager from a database.

4. The method as defined in claim 3, further comprising the step of:

passing marked search result items to a Graphical User Interface(GUI) / Event Manager.

5. The method as defined in claim 4, further comprising the step of:

associating an event handler to each search result item by a GUI / Event Manager.

6. The method as defined in claim 5, further comprising the step of:

displaying an enhanced search result item set in a display by a GUI / Event Manager.

7. The method as defined in claim 6, wherein the display comprises a web browser.

8. The method as defined in claim 7, further comprising the step of:

initiating a user selection process and notifying an event handler; and

receiving a user selection.

9. The method as defined in claim 8, wherein the user selection comprises a drag and drop mouse selection.

10. The method as defined in claim 8, further comprising the step of:

receiving notification in the GUI / Event Handler with selected source and target search result items.

11. The method as defined in claim 10, further comprising the step of:

forwarding selected source and target search result items to a downloader component.

12. The method as defined in claim 11, further comprising the step of:

attempting to access and retrieve search result documents represented by the selected source and target search result items by the downloader component.

13. The method as defined in claim 12, further comprising the steps of:

determining if retrieval is possible and if the retrieval is not possible then sending an error message to the GUI / Event Handler.

14. The method as defined in claim 12, further comprising the steps of:

determining if retrieval is possible and if the retrieval is possible then retrieving search result documents represented by the selected source and target search result items by the downloader component; and

forwarding the retrieved search result documents to a comparison unit.

15. The method as defined in claim 14, further comprising the steps of:

receiving retrieved search result documents in a comparison unit; and

beginning comparison of the retrieved search result documents.

16. The method as defined in claim 15, further comprising the steps of:

computing a similarity value for the retrieved search result documents; and

forwarding the value to the GUI / Event Manager.

17. The method as defined in claim 16, further comprising the step of:

generating a display graphic of the similarity value.

18. The method as defined in claim 17, wherein the display graphic comprises a Venn Diagram.

19. An information processing system for comparison of documents found on a network interconnected with a plurality of information processing units and hub processing units, the information processing system comprising:

a selection module for receiving a user selection request to select documents for comparison; and

a comparison module for comparing contents of at least two documents for similarity, wherein the comparison module returns a numeric similarity value which represents the similarity of the documents.

20. A computer readable program product for comparison of documents found on a network interconnected with a plurality of information processing units and

hub processing units, the computer readable program product comprising instructions for:

receiving a user selection request to select documents for comparison; and

comparing at least two documents for similarity, wherein the instruction of comparing documents includes returning a numeric similarity value which represents the similarity of the documents.

21. The computer readable program product as defined in claim 20, wherein the user selection request comprises a drag and drop mouse selection.

22. The computer readable program product as defined in claim 20, further comprising the instruction of:

computing a similarity percentage for the selected documents.

23. A method for comparison of documents found on a network interconnected with a plurality of information processing units and hub processing units, the method on a hub processing unit comprising the steps of:

receiving retrieved search result documents in a comparison unit; and

beginning comparison of the retrieved search result documents, wherein the comparison includes returning a numeric similarity value which represents the similarity of the documents.

24. The method as defined in claim 23, further comprising the steps of:

forwarding the value to a GUI / Event Manager.

25. The method as defined in claim 24, further comprising the step of:

generating a display graphic of the similarity value for forwarding to the GUI / Event Manager.

26. The method as defined in claim 25, wherein the display graphic comprises a Venn Diagram.